The data gathered for books was broken into five major categories: BestSeller, Rating, Review Count, Age Group, and Keywords. The BestSeller, Rating, and Review Count were chosen as indicators of popularity, fame, and quality. Age Group and Keywords were chosen as subject matter indicators. A total of forty-three binary ANN inputs are used to represent a book's data.

It was originally intended to acquire BestSeller data from the New York Times and Amazon, however, only data from Amazon ended up being acquired. BestSeller data is represented as a single binary input with a 1 representing that the book is an Amazon #1 BestSeller, and a 0 if not.

The Rating for a book is represented with four mutually-exclusive binary inputs. These four inputs represent ranges of Amazon star-ratings as indicated in the following table.

A book's Review Count is represented with fifteen mutually-exclusive binary inputs. These fifteen inputs represent ranges of a logarithmic scale as indicated in the following table. A book's total review count on Amazon indicates which category it falls in.

The Age Group for a book is represented with three mutually-exclusive binary inputs. These three inputs represent one of either Child, Teen, or Adult literature. Amazon's book description and book categorization is used to determine if a book belongs in Child or Teen literature, otherwise it is considered Adult literature.

Book Keywords are represented with twenty binary inputs. These twenty inputs represent any of the substrings from the following table that are present in Amazon's book description and book categorization.